

BUNDESREPUBLIK DEUTSCHLAND



Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 102 44 165.0

Anmeldetag: 23. September 2002

Anmelder/Inhaber: Infineon Technologies AG,
München/DE

Bezeichnung: Verfahren zur rechnergestützten Spracherkennung,
Spracherkennungssystem und Steuereinrichtung zum
Steuern eines technischen Systems und
Telekommunikationsgerät

IPC: G 10 L 15/02

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 15. Mai 2003
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag



wenner

Beschreibung

**Verfahren zur rechnergestützten Spracherkennung,
Spracherkennungssystem und Steuereinrichtung zum Steuern
5 eines technischen Systems und Telekommunikationsgerät**

Die Erfindung betrifft ein Verfahren zur rechnergestützten
Spracherkennung, ein Spracherkennungssystem sowie eine
Steuereinrichtung zum Steuern eines technischen Systems mit
10 einem Spracherkennungssystem und ein Telekommunikationsgerät.

Im Rahmen der rechnergestützten Spracherkennung wird ein von
einem Benutzer eingesprochenes Sprachsignal im Rahmen der
Vorverarbeitung digitalisiert und auf so genannte
15 Merkmalsvektoren, die auch als Featurevektoren bezeichnet
werden, abgebildet und für die durchzuführende
Spracherkennung gespeichert.

Die Merkmalsvektoren weisen je nach Anwendung eine fest
20 vorgegebene Anzahl von Merkmalsvektor-Komponenten auf, die
üblicherweise in dem Merkmalsvektor geordnet sind nach ihrer
Bedeutung im Rahmen der Spracherkennung, üblicherweise
geordnet nach Merkmalsvektor-Komponenten mit geringer
werdendem Informationsgehalt (kleiner werdender statistischer
25 Varianz).

Insbesondere in einer Spracherkennungsanwendung in einem
Embedded System ist jedoch die zur Verfügung stehende
Rechenleistung und der zur Verfügung stehende Speicherplatz
30 knapp, weshalb es in den derzeit bekannten
Spracherkennungsanwendungen insbesondere aufgrund einer sehr
hohen Anzahl von Merkmalsvektor-Komponenten, häufig zu
Problemen kommt.

35 Der Erfindung liegt das Problem zu Grunde, eine Möglichkeit
zur rechnergestützten Spracherkennung sowie ein
Spracherkennungssystem anzugeben, bei der eine verringerte

zur Verfügung stehenden Rechenleistung oder ein reduzierter
zur Verfügung stehenden Speicherplatz ausreicht.

Das Problem wird durch das Verfahren zur rechnergestützten
5 Spracherkennung, durch das Spracherkennungssystem, durch die
Steuereinrichtung sowie durch das Telekommunikationsgerät mit
den Merkmalen gemäß den unabhängigen Patentansprüchen gelöst.

Bei einem Verfahren zur rechnergestützten Spracherkennung
10 unter Verwendung von Merkmalsvektoren ist eine, vorzugsweise
zu Beginn des Verfahrens, ermittelte Erkennungsraten-
Information gespeichert, mit der für die Merkmalsvektoren
abhängig von dem Informationsgehalt der Merkmalsvektor-
Komponenten angegeben wird, welche Spracherkennungsrate
15 jeweils mit den Merkmalsvektoren mit den jeweils
berücksichtigten Merkmalsvektor-Komponenten erzielbar ist.

In einem ersten Schritt wird für eine
Spracherkennungsanwendung ermittelt oder bestimmt, welche
20 Spracherkennungsrate für die jeweilige
Spracherkennungsanwendung benötigt wird.

Unter Verwendung der gespeicherten Spracherkennungsraten-
Information wird von dem Rechner ermittelt, welcher
25 Informationsgehalt der Merkmalsvektor-Komponenten mindestens
erforderlich ist, um die bestimmte Spracherkennungsrate zu
gewährleisten.

Ferner wird ermittelt, wie viele Merkmalsvektor-Komponenten
30 in dem Spracherkennungssystem für die jeweilige
Spracherkennungsanwendung erforderlich sind, um den
ermittelten Informationsgehalt bereitzustellen.

Vorzugsweise wird ferner für die jeweilige
35 Spracherkennungsanwendung ein Codebuch erstellt unter
Berücksichtigung der zuvor ermittelten Anzahl von
Merkmalsvektor-Komponenten in dem Spracherkennungssystem.

Anschließend wird - vorzugsweise unter Verwendung des bestimmten, Spracherkennungsanwendungs-spezifischen Codebuchs - die Spracherkennung ausgeführt wird unter
5 Verwendung von Merkmalsvektoren mit der Anzahl von Merkmalsvektor-Komponenten, die erforderlich sind, um den ermittelten Informationsgehalt bereitzustellen.

Die Spracherkennung, das heißt das Verfahren zum Vergleichen
10 der Merkmalsvektoren, insbesondere somit der Vergleich der Merkmalsvektoren eines eingesprochenen Sprachsignals mit den Merkmalsvektoren von Referenzwörtern, die in einem elektronischen Wörterbuch gespeichert sind, wird ausgeführt unter Verwendung von Merkmalsvektoren mit der Anzahl von
15 Merkmalsvektor-Komponenten, die erforderlich ist, um die zuvor bestimmte Spracherkennungsrate zu gewährleisten.

Ein Spracherkennungssystem weist eine Spracherkennungseinheit auf sowie ein mit der Spracherkennungseinheit gekoppeltes
20 elektronisches Wörterbuch, in dem die im Rahmen der Spracherkennung berücksichtigten Wörter gespeichert sind. Ferner ist in dem Spracherkennungssystem ein Erkennungsraten- Informations-Speicher vorgesehen, in dem Erkennungsraten- Information gespeichert ist, mit der für die Merkmalsvektoren
25 abhängig von dem Informationsgehalt der Merkmalsvektor-Komponenten angegeben wird, welche Spracherkennungsrate jeweils mit den Merkmalsvektoren mit den jeweils berücksichtigten Merkmalsvektor-Komponenten erzielbar ist. Mittels einer ebenfalls vorgesehenen Erkennungsraten-
30 Informations-Ermittlungseinheit zum Ermitteln der Erkennungsraten-Information wird vor Durchführung der eigentlichen Spracherkennung anhand vorzugsweise eines Trainingsdatensatzes die Erkennungsraten-Information ermittelt. Ferner ist eine Informationsgehalt-
35 Ermittlungseinheit vorgesehen zum Ermitteln des Informationsgehalts für Merkmalsvektor-Komponenten eines Merkmalsvektors in dem Spracherkennungssystem. Ferner ist

eine Merkmalsvektor-Komponenten-Auswahleinheit zum Auswählen von Merkmalsvektor-Komponenten, die im Rahmen der Spracherkennung zu berücksichtigen sind, in dem Spracherkennungssystem vorgesehen.

5

Eine Steuereinrichtung zum Steuern eines technischen Systems weist das oben beschriebene Spracherkennungssystem auf, wobei in dem elektronischen Wörterbuch die zum Steuern des technischen Systems vorgesehenen Steuerbefehle zur, vorzugsweise sprecherunabhängigen, Spracherkennung gespeichert sind.

10

Anschaulich ist somit erfindungsgemäß erstmals ermöglicht, die tatsächlichen anwendungsspezifischen Anforderungen an die Erkennungsrate im Rahmen der Auswahl von Merkmalsvektor-Komponenten von Merkmalsvektoren zur Spracherkennung flexibel zu berücksichtigen, ohne dass für jede Spracherkennungsanwendung erneut eine Spracherkennungsrate ermittelt werden muss.

20

Auf diese Weise wird ein optimierter Kompromiss insbesondere hinsichtlich des zur Verfügung stehenden Speicherplatzbedarfs durch anwendungsabhängige Reduktion der Dimension der Merkmalsvektoren, anders ausgedrückt der Anzahl berücksichtigter Merkmalsvektor-Komponenten erreicht. Die Reduktion der Anzahl berücksichtigter Merkmalsvektor-Komponenten im Rahmen der Spracherkennung führt zu einer erheblichen Reduktion der im Rahmen der Spracherkennung selbst benötigten Rechnerleistung.

25

30

Aus diesem Grund eignet sich die Erfindung insbesondere für den Einsatz in einem Embedded System.

Ferner wird eine erhebliche Einsparung an benötigter Rechenzeit erreicht, da für eine neue Spracherkennungsanwendung lediglich die Anzahl erforderlicher Merkmalsvektor-Komponenten aus der zuvor lediglich einmal

35

ermittelten Erkennungsraten-Information bestimmt werden braucht und das Codebuch unmittelbar unter Verwendung der Merkmalsvektoren mit der bestimmten erforderlichen Anzahl von Merkmalsvektor-Komponenten ermittelt werden kann.

5

Bevorzugte Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Die im Folgenden beschriebenen Ausgestaltungen der Erfindung
10 betreffen sowohl das Verfahren, das Spracherkennungssystem als auch die Steuereinrichtung.

Für die Spracherkennung selbst wird vorzugsweise ein
Spracherkennungsverfahren zur sprecherunabhängigen
15 Spracherkennung, besonders bevorzugt unter Verwendung von Hidden Markov Modellen durchgeführt.

Alternativ können zur Spracherkennung, insbesondere zur
sprecherunabhängigen Spracherkennung statistische
20 Klassifikatoren, beispielsweise unter Verwendung künstlicher neuronaler Netze, eingesetzt werden.

Allgemein kann jedoch erfindungsgemäß jedes beliebige
Verfahren zur Spracherkennung eingesetzt werden.

25

Gemäß einer anderen Ausgestaltung der Erfindung ist es
vorgesehen, dass die Merkmalsvektor-Komponenten mit relativ
hohem Informationsgehalt unter den Merkmalsvektor-Komponenten
des jeweiligen Merkmalsvektors ausgewählt werden und im
30 Rahmen der Spracherkennung verwendet werden.

Durch diese Ausgestaltung der Erfindung wird gewährleistet,
dass tatsächlich diejenigen Merkmalsvektor-Komponenten nicht
berücksichtigt werden, die den geringsten Informationsgehalt
35 innerhalb aller Merkmalsvektor-Komponenten aufweisen, womit gewährleistet wird, dass die verloren gegangene Information im Rahmen der Spracherkennung, die entsteht aufgrund der

Nicht-Berücksichtigung einer Merkmalsvektor-Komponenten, minimiert ist.

Als Steuereinrichtung zum Steuern eines technischen Systems
5 eignen sich beispielsweise eine Steuereinrichtung zum Steuern
eines Telekommunikationsgeräts, beispielsweise eines
Telefongeräts, eines Telefaxgeräts, eines PDAs, eines
Notebooks, etc., oder zum Steuern eines Endgeräts, in dem
mindestens zwei der oben beschriebenen Geräte-
10 Funktionalitäten in einem gemeinsamen Gerät integriert sind.
Insbesondere diese mit einem klar definierten und begrenzten
Wortschatz zu steuernden Geräte können mittels eines
Sprachdialogs gesteuert werden, der relativ übersichtlich und
somit selbst mittels eines Embedded Systems kostengünstig
15 realisierbar ist.

Die anwendungsangepasste erhebliche Reduktion der Dimension
verarbeiteter Merkmalsvektoren führt zu einer erheblichen
Zeiteinsparung im Rahmen der Entwicklung eines
20 Spracherkennungssystems, insbesondere wird das verwendete
Codebuch erheblich reduziert, womit der Speicherplatzbedarf
ebenfalls in erheblichem Maße reduziert wird.

Ein Ausführungsbeispiel der Erfindung ist in den Figuren
25 dargestellt und wird im Folgenden näher erläutert.

Es zeigen

Figur 1 ein Blockdiagramm eines Spracherkennungssystems gemäß
30 einem Ausführungsbeispiel der Erfindung;

Figur 2 eine Skizze des Speichers des Rechners aus Figur 1 im
Detail;

35 Figur 3 ein Blockdiagramm, in dem die einzelnen
Verfahrensschritte zum Bestimmen einer

Erkennungsraten-Information gemäß einem Ausführungsbeispiel der Erfindung dargestellt sind;

Figur 4 ein Ablaufdiagramm, in dem die einzelnen
5 Verfahrensschritte zum Bestimmen einer
Erkennungsraten-Information gemäß einem
Ausführungsbeispiel der Erfindung dargestellt sind;

Figur 5 eine Skizze einer Erkennungsraten-Information gemäß
10 einem Ausführungsbeispiel der Erfindung;

Figur 6 ein Ablaufdiagramm, in dem die einzelnen
Verfahrensschritte des Verfahrens zur Spracherkennung
gemäß einem Ausführungsbeispiel der Erfindung
15 dargestellt sind.

Fig.1 zeigt ein Spracherkennungssystem 100 gemäß einem
Ausführungsbeispiel der Erfindung.

20 Das Spracherkennungssystem 100 arbeitet je nach Betriebsmodus
in einem ersten Betriebsmodus als
Spracherkennungseinrichtung, wobei in dem
Spracherkennungsmodus eine eingesprochene Äußerung 101,
eingesprochen von einem Benutzer (nicht dargestellt) des
25 Spracherkennungssystems 100, von der
Spracherkennungseinrichtung erkannt wird. Die Spracherkennung
erfolgt unter Verwendung eines Verfahrens zur
sprecherunabhängigen Spracherkennung.

30 In einem zweiten Betriebsmodus, im Weiteren auch bezeichnet
als Trainingsmodus, wird unter Verwendung einer
eingesprochenen Äußerung 101, wie im Weiteren näher erläutert
wird, das Spracherkennungssystem 100 trainiert, gemäß diesem
Ausführungsbeispiel bedeutet dies, dass einzelne Hidden
35 Markov Modelle für eine Äußerung mittels der eingesprochenen
Äußerung 101 trainiert werden.

In beiden Betriebsmodi wird das von dem Benutzer
eingesprochene Sprachsignal 101 einem Mikrofon 102 zugeführt
und als aufgenommenes elektrisches Analogsignal 103 einer
Vorverstärkung mittels einer Vorverstärkungseinheit 104
5 unterzogen und als verstärktes Analogsignal 105 einem Analog-
/Digitalwandler 106 zugeführt, dort in ein digitales Signal
107 umgewandelt und als digitales Signal 107 einer
Merkmalsextraktionseinheit 108 zugeführt, welche das digitale
Signal 107 einer Spektraltransformation unterzieht und zu dem
10 digitalen Signal 107 zu einer Äußerung eine Folge von
Merkmalsvektoren 109 bildet, welche das digitale Signal 107
repräsentieren.

Jeder Merkmalsvektor 109 weist eine vorgegebene Anzahl von
15 Merkmalsvektor-Komponenten auf.

Gemäß diesem Ausführungsbeispiel weisen die Merkmalsvektoren
jeweils 78 Merkmalsvektor-Komponenten auf.

20 Die Merkmalsvektoren 109 werden einem Rechner 110 zugeführt.

Es ist in diesem Zusammenhang anzumerken, dass das Mikrofon
102, die Vorverstärkungseinheit 104, insbesondere die
Verstärkungseinheit, und der Analog-/Digitalwandler 106 sowie
25 die Merkmalsextraktionseinheit 108 als separate Einheiten
oder auch als in dem Rechner 110 integrierte Einheiten
realisiert sein können.

Gemäß diesem Ausführungsbeispiel der Erfindung ist es
30 vorgesehen, dass die Merkmalsvektoren 109 dem Rechner 110
über dessen Eingangsschnittstelle 111 zugeführt werden.

Der Rechner 110 weist ferner einen Mikroprozessor 112, einen
Speicher 113 sowie eine Ausgangsschnittstelle 114 auf, welche
35 alle miteinander mittels eines Computerbus 115 gekoppelt
sind.

Mittels des Mikroprozessors 112 werden die im Folgenden beschriebenen Verfahrensschritte, insbesondere die Verfahren zum Ermitteln der im Folgenden erläuterten Erkennungsraten-Information sowie die Verfahren zur Spracherkennung
5 durchgeführt.

In einem im Folgenden näher erläuterten elektronischen Wörterbuch, welcher im Speicher 113 gespeichert ist, sind die Einträge in Form trainierter Hidden Markov Modelle enthalten,
10 die im Rahmen der Spracherkennung als Referenzwörter, die überhaupt nur von dem Spracherkennungsalgorithmus überhaupt erkannt werden können, enthalten sind.

Alternativ kann zusätzlich ein digitaler Signalprozessor
15 vorgesehen sein, der die jeweils eingesetzten Spracherkennungsalgorithmen implementiert hat und einen darauf spezialisierten Mikrocontroller aufweisen kann.

Ferner ist der Rechner 110 mittels der Eingangsschnittstelle
20 113 mit einer Tastatur 116 sowie einer Computermaus 117 über elektrische Leitungen 118, 119 oder eine Funkverbindung, beispielsweise eine Infrarot-Verbindung oder eine Bluetooth-Verbindung gekoppelt.

Über zusätzliche Kabel oder Funkverbindungen, beispielsweise
25 mittels einer Infrarot-Verbindung oder einer Bluetooth-Verbindung 120, 121 ist der Rechner 110 mittels der Ausgangsschnittstelle 114 mit einem Lautsprecher 122 sowie einem Aktor 123 gekoppelt.

30

Der Aktor 123 repräsentiert in Fig.1 allgemein jeden möglichen Aktor im Rahmen der Steuerung eines technischen Systems, beispielsweise realisiert in Form eines Hardwareschalters oder in Form eines Computerprogramms für
35 den Fall, dass beispielsweise ein Telekommunikationsgerät oder ein anderes technisches System, beispielsweise ein Autoradio, eine Stereoanlage, ein Videorekorder, ein

Fernseher, der Rechner 110 selbst oder irgendeine andere technische Anlage gesteuert werden soll.

Gemäß dem Ausführungsbeispiel der Erfindung weist die
5 Merkmalsextraktionseinheit 108 eine Filterbank mit einer
Mehrzahl von Bandpässen auf, welche die Energie des
einggegebenen Sprachsignals 103 in einzelnen Frequenzbändern
messen. Mittels der Filterbank werden so genannte
10 Kurzzeitspektren gebildet, indem die Ausgangssignale der
Bandpässe gleichgerichtet, geglättet und in kurzen Abständen
abgetastet werden, gemäß dem Ausführungsbeispiel alle
10 msec, alternativ alle 15 msec.

Die mittels der Merkmalsextraktionseinheit 108 gebildeten
15 Cepstrum-Koeffizienten, die 13 Koeffizienten der
Merkmalsvektoren 109 bilden, werden als Merkmalsvektor-
Komponenten von zwei aufeinander folgenden Zeitfenstern der
Größe von 10 msec oder von 15 msec in dem Merkmalsvektor 109
gespeichert. Ferner sind als Merkmalsvektor-Komponenten in
20 dem Merkmalsvektor 109 jeweils die zeitliche erste Ableitung
sowie die zeitliche zweite Ableitung der Cepstrum-
Koeffizienten in dem Merkmalsvektor 109 als Super-
Merkmalsvektor zusammengefasst und werden dem Rechner 110
zugeführt.

25 In dem Rechner 110 ist in Form eines Computerprogramms eine
Spracherkennungseinheit realisiert und in einem ersten
Speicherteilbereich 201 (vgl. **Fig.2**) in dem Speicher 113
gespeichert, welche Spracherkennungseinheit auf dem Prinzip
30 der Hidden Markov Modelle basiert. Somit erfolgt mittels des
Computerprogramms eine sprecherunabhängige Spracherkennung.

Zu Beginn des Verfahrens werden zwei unterschiedliche
Datensätze mit von einem oder mehreren Benutzern
35 eingesprochenen Sprachäußerungen gebildet.

Ein Trainingsdatensatz, gespeichert in einem zweiten Speicherteilbereich 202 des Speichers 113 weist diejenigen Sprachäußerungen, in Form von für die jeweiligen Sprachäußerungen gebildeten Merkmalsvektoren, auf, die zum im
5 Folgenden näher erläuterten Trainieren der Hidden Markov Modelle, welche zur Spracherkennung eingesetzt werden, verwendet werden.

10 In einem dritten Speicherteilbereich 203 ist ein Testdatensatz gespeichert, das heißt die Sprachäußerungen, die verwendet werden zum Testen der trainierten Spracherkennungseinheit, anders ausgedrückt zum Testen der trainierten Hidden Markov Modelle, die in einem vierten Speicherteilbereich 204 gespeichert sind.

15 Mittels des Testdatensatzes wird, wie im Folgenden näher erläutert wird, eine Erkennungsraten-Information ermittelt, welche in einem fünften Speicherteilbereich 205 gespeichert sind.

20 In einem sechsten Speicherteilbereich 206 ist ferner eine im Weiteren näher erläuterte Tabelle gespeichert, in der für eine oder mehrere Anwendungen des Spracherkennungssystems eine Angabe darüber gespeichert ist, welche Erkennungsrate
25 für die jeweilige Anwendung benötigt wird.

Es ist in diesem Zusammenhang darauf hinzuweisen, dass die einzelnen Elemente in unterschiedlichen Speicherbereichen desselben Speichers 113 gespeichert sein können, jedoch auch
30 in unterschiedlichen, vorzugsweise an die jeweiligen Anforderungen der gespeicherten Elemente angepassten Speichern.

Fig.3 und **Fig.4** zeigen in einem Blockdiagramm 300
35 (vgl. Fig.3) bzw. in einem Ablaufdiagramm (vgl. Fig.4) die einzelnen von dem Rechner 110 durchgeführten Verfahrensschritte des Verfahrens zum Ermitteln der in dem

fünften Speicherteilbereich 205 gespeicherten
Erkennungsraten-Information.

Nach Starten des Verfahrens (Schritt 401) werden in einem
5 Trainingsschritt die einzelnen Hidden Markov Modelle unter
Verwendung des in dem zweiten Teilspeicherbereich 202
gespeicherten Trainingsdatensatzes trainiert.

Das Training der Hidden Markov Modelle erfolgt gemäß diesem
10 Ausführungsbeispiel in drei Phasen:

- einer ersten Phase (Schritt 402), in der die in der
Trainings-Datenbank enthaltenen Sprachsignale 301
segmentiert werden mittels einer Segmentierungseinheit
302,
- 15 • einer zweiten Phase (Schritt 403), in der die LDA-Matrix
(lineare Diskriminanzanalyse-Matrix) berechnet wird
sowie
- einer dritten Phase (Schritt 405), in der das Codebuch,
das heißt die HMM-Prototypen-Merkmalvektoren für
20 jeweils eine in einem Auswahlsschritt (Schritt 404)
ausgewählte Anzahl von Merkmalsvektor-Komponenten
berechnet werden.

Die Gesamtheit dieser drei Phasen wird im Weiteren als das
25 Training der Hidden Markov Modelle bezeichnet (HMM-Training).

Das HMM-Training wird unter Verwendung des DSPs 123 sowie
unter Verwendung von vorgegebenen Trainingskripts,
anschaulich von geeignet eingerichteten Computerprogrammen,
30 durchgeführt.

Gemäß diesem Ausführungsbeispiel wird jede gebildete
lautsprachliche Einheit, das heißt jedes Phonem, in drei
aufeinander folgende Phonemsegmente aufgeteilt, entsprechend
35 einer Initial-Phase (erstes Phonemsegment), einer zentralen
Phase (zweites Phonemsegment) und einer Endphase (drittes
Phonemsegment) eines Lauts, das heißt eines Phonems.

Anders ausgedrückt wird jeder Laut in einem Lautmodell mit drei Zuständen, das heißt mit einem Drei-Zustands-HMM modelliert.

5

Während der Spracherkennung werden die drei Phonemsegmente in einer Bakis-Topologie oder allgemein einer Links-Rechts-Topologie aneinander gereiht und auf die Konkatenation dieser drei aneinander gereihten Segmente wird die Berechnung im Rahmen der sprecherunabhängigen Spracherkennung durchgeführt.

10

Wie im Weiteren noch näher erläutert wird, wird in dem Spracherkennungsmodus ein Viterbi-Algorithmus zum Dekodieren der Merkmalsvektoren, welche aus dem eingegebenen Sprachsignal 101 gebildet werden, durchgeführt.

15

Nach erfolgter Segmentierung wird die LDA-Matrix 304 (Schritt 403) mittels einer LDA-Matrix-Berechnungseinheit 303 ermittelt.

20

Die LDA-Matrix 304 dient zur Transformation eines jeweiligen Super-Merkmalsvektors \underline{y} auf einen Merkmalsvektor \underline{x} gemäß folgender Vorschrift:

$$\underline{x} = \underline{A}^T \cdot (\underline{y} - \underline{\bar{y}}), \quad (1)$$

25

wobei mit

- \underline{x} ein Merkmalsvektor,
- \underline{A} eine LDA-Matrix,
- \underline{y} ein Super-Merkmalsvektor,
- $\underline{\bar{y}}$ ein globaler Verschiebungsvektor

30

bezeichnet wird.

35

Die LDA-Matrix \underline{A} wird derart bestimmt, dass

- die Komponenten des Merkmalsvektors \underline{x} im statistischen Durchschnitt voneinander im Wesentlichen unkorreliert sind,
- die statistischen Varianzen innerhalb einer Segmentklasse im statistischen Durchschnitt normalisiert sind,
- die Zentren der Segmentklassen im statistischen Durchschnitt einen maximalen Abstand voneinander aufweisen und
- die Dimension der Merkmalsvektoren \underline{x} möglichst, vorzugsweise Spracherkennungsanwendungs-abhängig, reduziert wird.

Im Folgenden wird das Verfahren zum Bestimmen der LDA-Matrix \underline{A} gemäß diesen Ausführungsbeispielen erläutert.

Es ist jedoch anzumerken, dass alternativ alle bekannten Verfahren zum Bestimmen einer LDA-Matrix \underline{A} ohne Einschränkung eingesetzt werden kann.

Es wird angenommen, dass J Segmentklassen existieren, wobei jede Segmentklasse j einen Satz D_y -dimensionaler Super-Merkmalvektoren \underline{y} enthält, das heißt, dass gilt:

$$\text{Klasse } j = \left\{ \underline{y}_j^1, \underline{y}_j^2, \dots, \underline{y}_j^{N_j} \right\}, \quad (2)$$

wobei mit N_j die Anzahl der in der Klasse j sich befindenden Super-Merkmalvektoren \underline{y}_j bezeichnet wird.

Mit

$$N = \sum_{j=1}^J N_j \quad (3)$$

wird die Gesamtzahl der Super-Merkmalvektoren \underline{y} bezeichnet.

Es ist anzumerken, dass die Super-Merkmalvektoren \underline{y}_j^k unter Verwendung der oben beschriebenen Segmentierung der Sprachsignal-Datenbank ermittelt worden sind.

5

Gemäß diesem Ausführungsbeispiel weist jeder Super-Merkmalvektor \underline{y}_j^k eine Dimension D_Y von

$$D_Y = 78 \quad (= 2 \cdot 3 \cdot 13)$$

10

auf, wobei 13 MFCC-Koeffizienten (Cepstrums-Koeffizienten) in dem Super-Merkmalvektor \underline{y}_j^k enthalten sind, sowie deren jeweilige zeitliche erste Ableitung und deren jeweilige zeitliche zweite Ableitung (dies begründet obigen Faktor 3).

15

Ferner sind in jedem Super-Merkmalvektor \underline{y}_j^k jeweils die Komponenten zweier zeitlich unmittelbar aufeinanderfolgender Zeitfenster im Rahmen der Kurzzeitanalyse enthalten (dies begründet obigen Faktor 2).

20

Es ist in diesem Zusammenhang anzumerken, dass grundsätzlich eine beliebige, an die jeweilige Anwendung angepasste Zahl von Vektorkomponenten in dem Super-Merkmalvektor \underline{y}_j^k enthalten sein kann, beispielsweise bis zu 20 Cepstrums-Koeffizienten und deren zugehörigen zeitlichen erste Ableitungen und zweite Ableitungen.

25

Der statistische Mittelwert oder anders ausgedrückt das Zentrum der Klasse j ergibt sich gemäß folgender Vorschrift:

30

$$\bar{\underline{y}}_j = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} \underline{y}_j^i. \quad (4)$$

Die Kovarianzmatrix $\underline{\Sigma}_j$ der Klasse j ergibt sich gemäß folgender Vorschrift:

$$\underline{\Sigma}_j = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} \left(\underline{y}_j^i - \bar{\underline{y}}_j \right) \cdot \left(\underline{y}_j^i - \bar{\underline{y}}_j \right)^T. \quad (5)$$

5

Die Durchschnitts-Intra-Streumatrix \underline{S}_w ist definiert als:

$$\underline{S}_w = \sum_{j=1}^J p(j) \cdot \underline{\Sigma}_j, \quad (6)$$

10 mit

$$p(j) = \frac{N_j}{N}, \quad (7)$$

wobei $p(j)$ als Gewichtungsfaktor der Klasse j bezeichnet wird.

15

In analoger Weise ist die Durchschnitts-Inter-Streumatrix \underline{S}_b definiert als:

$$\underline{S}_b = \sum_{j=1}^J p(j) \cdot \left(\bar{\underline{y}}_j - \bar{\underline{y}} \right) \cdot \left(\bar{\underline{y}}_j - \bar{\underline{y}} \right)^T, \quad (8)$$

mit

$$\bar{\underline{y}} = \sum_{j=1}^J p(j) \cdot \bar{\underline{y}}_j \quad (9)$$

25

als dem Durchschnitts-Super-Merkmalvektor über alle Klassen.

Die LDA-Matrix \underline{A} wird zerlegt gemäß folgender Vorschrift:

$$\underline{A} = \underline{U} \cdot \underline{W} \cdot \underline{V}, \quad (10)$$

30

wobei mit

- \underline{U} eine erste Transformationsmatrix,
- 5 • \underline{W} eine zweite Transformationsmatrix und
- \underline{V} eine dritte Transformationsmatrix

bezeichnet wird.

- 10 Die erste Transformationsmatrix \underline{U} wird verwendet, um die Durchschnitts-Intra-Streumatrix \underline{S}_W zu diagonalisieren und wird ermittelt, indem die positiv definite und symmetrische Durchschnitts-Intra-Streumatrix \underline{S}_W in ihren Eigenvektorraum transformiert wird. In ihrem Eigenvektorraum ist die
- 15 Durchschnitts-Intra-Streumatrix \underline{S}_W eine Diagonal-Matrix, deren Komponenten positiv und größer oder gleich null sind. Die Komponenten, deren Werte größer null sind, entsprechen der Durchschnitts-Varianz in der jeweiligen durch die entsprechende Vektorkomponente definierten Dimension.

20

Die zweite Transformationsmatrix \underline{W} wird zum Normalisieren der Durchschnitts-Varianzen verwendet und wird ermittelt gemäß folgender Vorschrift:

25
$$\underline{W} = \left(\underline{U}^T \cdot \underline{S}_W \cdot \underline{U} \right)^{-\frac{1}{2}} . \quad (11)$$

Die Transformation $\underline{U} \cdot \underline{W}$ wird auch als Weißung bezeichnet.

Mit

30

$$\underline{B} = \underline{U} \cdot \underline{W} \quad (12)$$

ergibt sich für die Matrix $\underline{B}^T \cdot \underline{S}_W \cdot \underline{B}$ die Einheitsmatrix, welche bei jeder beliebigen orthonormalen

- 35 Lineartransformation unverändert bleibt.

Um die Durchschnitts-Inter-Streumatrix \underline{S}_b zu diagonalisieren wird die dritte Transformationsmatrix \underline{V} , die gebildet wird gemäß folgender Vorschrift:

5

$$\underline{V} = \underline{B}^T \cdot \underline{S}_b \cdot \underline{B}, \quad (13)$$

wobei $\underline{B}^T \cdot \underline{S}_b \cdot \underline{B}$ ebenfalls eine positiv definite und symmetrische Matrix darstellt, in ihren Eigenvektorraum
10 transformiert wird.

In dem Transformationsraum

$$\underline{x} = \underline{A}^T \cdot (\underline{y} - \bar{\underline{y}}) \quad (14)$$

15

ergeben sich somit folgende Matrizen:

Eine diagonalisierte Durchschnitts-Intra-Streumatrix \underline{S}_w :

$$\underline{S}_w = \text{diag}(\underline{1})_{d=1 \dots D_Y} \quad (15)$$

und eine diagonalisierte Durchschnitts-Inter-Streumatrix \underline{S}_b :

$$\underline{S}_b = \text{diag}(\sigma_d^2)_{d=1 \dots D_Y}, \quad (16)$$

25

wobei mit $\text{diag}(c_d)_{d=1 \dots D_Y}$ eine $D_Y \times D_Y$ Diagonalmatrix mit den Komponenten c_d in der Zeile/Spalte d und sonst mit Komponenten mit dem Wert Null, bezeichnet wird.

30 Die Werte σ_d^2 sind die Eigenwerte der Durchschnitts-Inter-Streumatrix \underline{S}_b und stellen ein Maß für die so genannte Pseudoentropie der Merkmalsvektor-Komponenten dar, welche im Folgenden auch als Informationsgehalt der Merkmalsvektor-Komponenten bezeichnet wird. Es ist anzumerken, dass die Spur

jeder Matrix invariant ist bezüglich irgendeiner Orthogonaltransformation, womit sich ergibt, dass die Summe

$$\sigma^2 = \sum_{d=1}^{D_Y} \sigma_d^2 \quad (17)$$

5

die Gesamt-Durchschnitts-Varianz des Durchschnitts-Vektors \underline{x}_j der J Klassen darstellt.

10

Es ergibt sich somit eine ermittelte Anhängigkeit der Pseudoentropie der Merkmalsvektoren von den jeweils in dem Merkmalsvektor enthaltenen bzw. berücksichtigten Merkmalsvektor-Komponenten.

15

Gemäß diesem Ausführungsbeispiel wird anschließend eine Dimensionsreduktion vorgenommen, indem die σ_d^2 -Werte in in ihrer Größe abfallender Reihenfolge sortiert werden und die σ_d^2 -Werte weggelassen werden, das heißt unberücksichtigt bleiben, die kleiner sind als ein vorgegebener Schwellwert. Der vorgegebene Schwellwert kann ferner kumulativ definiert sein.

20

Dann kann die LDA-Matrix \underline{A} angepasst werden, indem die Zeilen entsprechend den Eigenwerten σ_d^2 sortiert werden und die Zeilen weggelassen werden, die zu den ausreichend „kleinen“ Varianzen gehören und damit nur einen geringen Informationsgehalt (geringe Pseudoentropie) aufweisen.

25

Gemäß diesem Ausführungsbeispiel werden die Komponenten mit den 24 größten Eigenwerten σ_d^2 verwendet, anders ausgedrückt

30

$D_X = 24$.

Die vier oben beschriebenen Teilschritte zum Ermitteln der LDA-Matrix \underline{A} 304 (Schritt 403) sind in folgender Tabelle zusammengefasst:

Nummer Verfahrensschritt	Ziel	Verfahren
1	Dekorrelieren der Merkmalsvektorkomponenten	Diagonalisieren der Durchschnitts-Intra-Klassen-Kovarianzmatrix \underline{S}_w
2	Normalisieren der statistischen Varianzen innerhalb einer Klasse	Bestimmen der inversen Quadratwurzel der transformierten Durchschnitts-Intra-Klassen-Kovarianzmatrix $\underline{U}^T \cdot \underline{S}_w \cdot \underline{U}$
3	Maximieren der Klassenzentren	Diagonalisieren der transformierten Durchschnitts-Inter-Klassen-Kovarianzmatrix $\underline{B}^T \cdot \underline{S}_b \cdot \underline{B}$
4	Reduzieren der Dimensionen der Merkmalsvektoren	Auswählen der Zeilen der Matrix \underline{A} mit den 24 größten Eigenwerten von $\underline{A}^T \cdot \underline{S}_b \cdot \underline{A}$

Das letzte Verfahren zum Teil-Verfahren im Rahmen des Trainings der Hidden Markov Modelle ist das Clustern der Merkmalsvektoren (Schritt 405), welches mittels einer Clustereinheit 305 durchgeführt wird und welches als Ergebnis ein jeweiliges Codebuch 306 hat, jeweils spezifisch für einen Trainingsdatensatz mit einer vorgegebenen Anzahl von Merkmalsvektor-Komponenten.

Die Gesamtheit der Repräsentanten der Segmentklassen wird als Codebuch bezeichnet und die Repräsentanten selbst werden auch als Prototypen der Phonemsegmentklasse bezeichnet.

5

Die Prototypen, im Weiteren auch als Prototyp-Merkmalsvektoren bezeichnet, werden gemäß dem in [1] beschriebenen Baum-Welch-Training ermittelt.

10 Somit sind die Basiseinträge des elektronischen Wörterbuches, das heißt die Basiseinträge zur sprecherunabhängigen Spracherkennung erstellt und gespeichert und die entsprechenden Hidden Markov Modelle trainiert.

15 Somit existiert für jeden Basiseintrag jeweils ein Hidden Markov Modell, womit das Codebuch 306 für den Trainingsdatensatz mit der ausgewählten Anzahl von Merkmalsvektor-Komponenten in den Merkmalsvektoren in dem Trainingsdatensatz.

20

Nach erfolgreichem Training der Hidden Markov Modelle liegen nunmehr die trainierten Hidden Markov Modelle in dem vierten Speicherteilbereich 204 vor.

25 In einem anschließenden Verfahrensschritt (Schritt 406) wird für die in dem Testdatensatz, welcher in dem dritten Teilspeicherbereich 203 gespeichert ist, die Erkennungsrate für die jeweiligen Merkmalsvektoren der aktuellen Dimension, das heißt für die Merkmalsvektoren mit der jeweils aktuellen
30 Anzahl von Merkmalsvektor-Komponenten, ermittelt.

Dies erfolgt gemäß diesem Ausführungsbeispiel dadurch, dass für alle Sprachäußerungen, das heißt für alle Folgen von Merkmalsvektoren in dem Testdatensatz eine Spracherkennung
35 mittels der trainierten Hidden Markov Modelle, anders ausgedrückt mittels einer Spracherkennungseinheit 307,

durchgeführt wird und die Spracherkennungsergebnisse mit den Soll-Ergebnissen des Testdatensatzes verglichen werden.

Die ermittelte Erkennungsrate 308 ergibt sich aus dem
5 Verhältnis der Anzahl korrekter Erkennungsergebnisse, anders
ausgedrückt aus der Anzahl von Übereinstimmungen zwischen dem
Spracherkennungsergebnis und dem Soll-Ergebnis, welches in
dem Testdatensatz angegeben ist, und der insgesamt zur
Spracherkennung dargestellten Testdatensätze.

10

In einem nachfolgenden Schritt (Schritt 304) wird die
ermittelte Erkennungsrate gemeinsam mit der Angabe, wie viele
Merkmalsvektor-Komponenten zur Bestimmung der Erkennungsrate
308 für die Merkmalsvektoren des Testdatensatzes 203

15 verwendet worden sind, gespeichert.

Anschließend wird in einem Prüfschritt 407 überprüft, ob das
Verfahren beendet werden soll.

20 Ist dies der Fall, so wird das Verfahren beendet.
(Schritt 408).

Soll das Verfahren noch nicht beendet werden, so wird die
Anzahl der Merkmalsvektor-Komponenten der Merkmalsvektoren
25 109, die im Rahmen der Ermittlung der Erkennungsrate aus dem
Testdatensatz verwendet werden, um einen vorgegebenen Wert,
vorzugsweise um den Wert „1“, das heißt um eine
Merkmalsvektor-Komponente reduziert (Schritt 409).

30 Anschließend werden die Schritte des Clusters (Schritt 405)
und somit des Erstellens des jeweiligen Codebuchs 306 und des
Bestimmens der Spracherkennungsrate (Schritt 406) erneut
durchgeführt, nunmehr jedoch für Merkmalsvektoren des
Testdatensatzes mit jeweils um eine Merkmalsvektor-Komponente
35 reduziertem Merkmalsvektoren.

Anders ausgedrückt bedeutet dies, dass bei 78 Merkmalsvektor-Komponenten in einem üblichen Merkmalsvektor gemäß diesem Ausführungsbeispiel der Erfindung in der zweiten Iteration die Erkennungsrate für einen Merkmalsvektor mit 77

- 5 Merkmalsvektor-Komponenten durchgeführt wird, in der dritten Iteration mit 76 Merkmalsvektor-Komponenten, usw.

Gemäß einer alternativen Ausgestaltung der Erfindung ist es vorgesehen, unmittelbar nicht mit allen Merkmalsvektor-

- 10 Komponenten des Super-Merkmalsvektors (d.h. nicht mit allen 78 Merkmalsvektor-Komponenten), zu beginnen, sondern schon zu Beginn eine um einen anwendungsabhängigen Wert reduzierte Anzahl von Merkmalsvektor-Komponenten.

- 15 Ferner kann in jeder Iteration die Anzahl von Merkmalsvektor-Komponenten um mehr als um den Wert „1“ reduziert werden.

Somit liegen als Ergebnis diese oben beschriebenen Verfahrens einerseits eine Pseudoentropie-Abbildung und andererseits

- 20 eine Erkennungsraten-Abbildung vor.

Mit der Pseudoentropie-Abbildung wird eine Abhängigkeit der Pseudoentropie der Merkmalsvektoren von den berücksichtigten Merkmalsvektor-Komponenten angegeben, also eine Abhängigkeit des Informationsgehalts, auch als Informationsmaß bezeichnet, von den berücksichtigten Merkmalsvektor-Komponenten.

- 25 Mit der Erkennungsraten-Abbildung wird eine Abhängigkeit der Spracherkennungsrate der Merkmalsvektoren von den berücksichtigten Merkmalsvektor-Komponenten angegeben.

- 30 Aus der Pseudoentropie-Abbildung und der Erkennungsraten-Abbildung wird die Erkennungsraten-Information gebildet, indem eine Abhängigkeit der Spracherkennungsrate von der Pseudoentropie ermittelt wird unter Verwendung der jeweiligen berücksichtigten Merkmalsvektor-Komponenten. Es ist anzumerken, dass die Erkennungsraten-Information nunmehr

unabhängig ist von der Anzahl der berücksichtigten Merkmalsvektor-Komponenten.

Die Erkennungsraten-Information wird in dem fünften
5 Teilspeicherbereich 205 gespeichert.

Ergebnis dieses Verfahrens ist somit die in **Fig.5** in einem Funktionsdiagramm dargestellte Erkennungsraten-Information 500, die über einer ersten Achse, auf der die ermittelte
10 Pseudoentropie 501 aufgetragen ist, die erreichte Erkennungsrate 502 in Form von Daten-Tupeln 503 angibt.

Die Erkennungsraten-Information 500 stellt somit den Zusammenhang dar zwischen der Pseudoentropie und der mittels
15 des Spracherkennungssystems erzielbaren Erkennungsrate.

Es ist in diesem Zusammenhang darauf hinzuweisen, dass die Erkennungsraten-Information 500 nur einmal für jedes Spracherkennungssystem, das heißt für jeden trainierten Satz
20 von Hidden Markov Modellen durchgeführt werden muss.

Fig.6 zeigt in einem Ablaufdiagramm 600 die einzelnen Verfahrensschritte des Verfahrens zur Spracherkennung gemäß dem Ausführungsbeispiel der Erfindung.

25 Nach Starten des Verfahrens (Schritt 601) wird die Spracherkennungsanwendung ausgewählt oder bestimmt, in deren Rahmen die Spracherkennung durchgeführt werden soll (Schritt 602).

30 Als mögliche Anwendungen für die Spracherkennung sind gemäß diesem Ausführungsbeispiel folgende Spracherkennungsapplikationen vorgesehen:

- ein Sprachdialogsystem:
35 für ein Sprachdialogsystem mit einer Spracherkennungsrate von 92 - 93 % zu gewährleisten;

- ein Fahrzeug-Navigationssystem:
für diese Spracherkennungsapplikation ist eine
Spracherkennungsrate von ungefähr 95 % zu gewährleisten;
- eine Steuerung eines technischen Systems, gemäß dem
5 Ausführungsbeispiel eines Videorekorders:
für diese Spracherkennungsapplikation ist eine
Spracherkennung von ungefähr 95 % zu gewährleisten;
- eine Telefon-Anwendung:
für diese Anwendung ist eine Spracherkennungsrate von
10 95 % zu gewährleisten;
- ein Diktat, anders ausgedrückt das Erkennen von
Sprachinformation und Umsetzen des erkannten
Sprachsignals in ein Textverarbeitungssystem:
für diese Applikation ist die mit dem
15 Spracherkennungssystem maximal erreichbare
Spracherkennungsrate erforderlich, das heißt in diesem
Fall ist keine Reduktion von Merkmalsvektor-Komponenten
sinnvoll.

20 Für die jeweilige Spracherkennungsanwendung erfolgt unter
einem ebenfalls in dem zweiten Speicherteilbereich 202
gespeicherten, vorzugsweise Spracherkennungsanwendungs-
abhängigen Trainingsdatensatz eine Segmentierung der Super-
Merkmalsvektoren (Schritt 603) in der gleichen, oben
25 beschriebenen Weise.

Anschließend wird, ebenfalls in der gleichen, oben
beschriebenen Weise eine LDA-Berechnung durchgeführt
(Schritt 604), womit eine Spracherkennungsanwendungs-
30 abhängige LDA-Matrix 605 ermittelt wird.

Unter Verwendung der Spracherkennungsanwendungs-abhängigen
LDA-Matrix 605 wird eine Spracherkennungsanwendungs-abhängige
Pseudoentropie-Abbildung ermittelt, die einen Zusammenhang
35 darstellt zwischen der erreichbaren Pseudoentropie und der
jeweils berücksichtigten Anzahl von Merkmalsvektor-
Komponenten in den Merkmalsvektoren.

Die jeweilige Spracherkennungsanwendungs-abhängige Pseudoentropie-Abbildung wird in dem sechsten Speicherteilbereich 206 gespeichert.

5

Unter Verwendung der zuvor ermittelten benötigten Spracherkennungsrate und der in dem sechsten Speicherteilbereich 206 gespeicherten Erkennungsraten-Information wird für die ausgewählte Anwendung in einem

10 zusätzlichen Schritt die erforderliche Pseudoentropie ermittelt (Schritt 606).

Unter Verwendung der Spracherkennungsanwendungs-abhängigen Pseudoentropie-Abbildung, wie sie zuvor ermittelt worden ist,

15 wird in einem anschließenden Schritt (Schritt 607) ermittelt, wie viele Merkmalsvektor-Komponenten und welche Merkmalsvektor-Komponenten, gemäß diesem Ausführungsbeispiel die jeweils die Merkmalsvektor-Komponenten mit jeweils kleinstem Informationsgehalt, im Rahmen der Spracherkennung

20 weggelassen werden können, anders ausgedrückt unberücksichtigt bleiben können.

Ist in dem Schritt 607 nunmehr die Anzahl benötigter Merkmalsvektor-Komponenten für die ausgewählte Anwendung

25 ermittelt, so wird in einem nachfolgenden Schritt für die jeweilige Anwendung und für die bestimmte Anzahl von Merkmalsvektor-Komponenten ein Clustering durchgeführt (Schritt 608). Ergebnis des Clusterings ist ein Spracherkennungsanwendungs-abhängiges Codebuch 609, anders

30 ausgedrückt eine Menge Spracherkennungsanwendungs-abhängiger trainierter Hidden Markov Modelle, welches ebenfalls in dem Speicher gespeichert wird. Das Clusterverfahren ist gleich dem oben beschriebenen Clusterverfahren (Schritt 405) zum Bestimmen der Erkennungsraten-Information 500.

35

Anschließend erfolgt die sprecherunabhängige Spracherkennung unter Verwendung des gespeicherten

Spracherkennungsanwendungs-abhängigen Codebuchs 609
(Schritt 610).

Anders ausgedrückt bedeutet dies, dass eine anschließend
5 eingesprochene Äußerung eines Benutzers unter Verwendung der
Hidden Markov Modelle gemäß dem [1] beschriebenen Verfahren
zur sprecherunabhängigen Spracherkennung unter Verwendung des
Viterbi-Algorithmus durchgeführt wird (Schritt 610).

10 Wie zuvor beschrieben werden im Rahmen der Spracherkennung
die reduzierten Merkmalsvektoren berücksichtigt, das heißt
die Merkmalsvektoren ohne die nicht berücksichtigten
Merkmalsvektor-Komponenten.

15 Anders ausgedrückt bedeutet dies, dass bei k Merkmalsvektor-
Komponenten in einem Merkmalsvektor und bei n nicht
berücksichtigten Merkmalsvektor-Komponenten ($n < k$) lediglich
($k - n$) Merkmalsvektor-Komponenten im Rahmen der
Spracherkennung berücksichtigt werden müssen.

20 Somit findet auch der Vergleich in einem Vergleichsraum der
Dimension ($k - n$) statt.

Ferner wird erfindungsgemäß die Erkennungsraten-Information
25 nur einmal bestimmt; für jede neue Spracherkennungsanwendung
ist es lediglich erforderlich, unter Verwendung der
Erkennungsraten-Information 500 zu ermitteln, wie viele und
vorzugsweise welche Merkmalsvektor-Komponenten für die neue
Spracherkennungsanwendung erforderlich sind, und das Codebuch
30 für die ermittelte Anzahl benötigter Merkmalsvektor-
Komponenten zu bestimmen.

Fig.5 zeigt das Beispiel, dass für die ausgewählte Anwendung
eine Spracherkennungsrate von 95 % benötigt wird, in Fig.5
35 dargestellt mittels einer Schnittlinie 504.

Oberhalb der Schnittlinie befindende Datenpunkte repräsentieren eine Pseudoentropie, die größer ist als es eigentlich erforderlich wäre für die Anforderung der ausgewählten Anwendung, anders ausgedrückt, um eine
5 Erkennungsrate von 95 % zu gewährleisten.

Gemäß diesem Ausführungsbeispiel können zwei Merkmalsvektor-Komponenten weggelassen werden, womit die Dimension der verarbeiteten Merkmalsvektoren um den Wert 2 reduziert werden
10 konnte.

Anschaulich kann die Erfindung darin gesehen werden, dass für eine spezielle ausgewählte Spracherkennungsanwendung, beispielsweise aus dem Bereich Command and Control, anders
15 ausgedrückt für eine Steuereinrichtung, unter bestimmten Bedingungen eine geringere Erkennrate des Spracherkenners akzeptiert werden kann und diese Erkenntnis erfindungsgemäß umgesetzt wird in ein Reduzieren der Dimension der verarbeiteten Merkmalsvektoren.

20

Nach erfolgter Spracherkennung in Schritt 610 wird das Verfahren beendet (Schritt 611).

In diesem Dokument ist folgende Veröffentlichung zitiert:

- [1] E.G. Schukat-Talamazzini, Automatischer Spracherkennung,
Grundlagen, statistische Modelle und effiziente
5 Algorithmen, Vieweg Verlag, ISBN 3-528-05492-1,
Seite 121 - 164, 1995

Patentansprüche

1. Verfahren zur rechnergestützten Spracherkennung unter Verwendung von Merkmalsvektoren, wobei eine Erkennungsraten-
5 Information gespeichert ist, mit der für die Merkmalsvektoren abhängig von dem Informationsgehalt der Merkmalsvektor-Komponenten angegeben wird, welche Spracherkennungsrate jeweils mit den Merkmalsvektoren mit den jeweils berücksichtigten Merkmalsvektor-Komponenten erzielbar ist,
10 • bei dem bestimmt wird, welche Spracherkennungsrate für eine Spracherkennungsanwendung benötigt wird,
• bei dem unter Verwendung der Erkennungsraten-Information ermittelt wird, welcher Informationsgehalt der Merkmalsvektor-Komponenten mindestens erforderlich ist,
15 um die bestimmte Spracherkennungsrate zu gewährleisten,
• bei dem ermittelt wird, wie viele Merkmalsvektor-Komponenten in dem Spracherkennungssystem für die Spracherkennungsanwendung erforderlich sind, um den ermittelten Informationsgehalt bereitzustellen,
20 • bei dem die Spracherkennung ausgeführt wird unter Verwendung von Merkmalsvektoren mit der Anzahl von Merkmalsvektor-Komponenten, die erforderlich sind, um den ermittelten Informationsgehalt bereitzustellen.
- 25 2. Verfahren gemäß Anspruch 1, bei dem für die Spracherkennung ein sprecherunabhängiges Spracherkennungsverfahren verwendet wird.
3. Verfahren gemäß Anspruch 2,
30 bei dem die Spracherkennung unter Verwendung von Hidden Markov Modellen durchgeführt wird.
4. Verfahren gemäß einem der Ansprüche 1 bis 3, bei dem die Merkmalsvektor-Komponenten mit höchstem
35 Informationsgehalt ausgewählt werden und im Rahmen der Spracherkennung verwendet werden.

5. Spracherkennungssystem mit

- einer Spracherkennungseinheit,
- einem mit der Spracherkennungseinheit gekoppelten elektronischen Wörterbuch, in dem die im Rahmen der Spracherkennung berücksichtigten Wörter gespeichert sind,
- einem Erkennungsraten-Informationen-Speicher, in dem Erkennungsraten-Information gespeichert ist, mit der für die Merkmalsvektoren abhängig von dem Informationsgehalt der Merkmalsvektor-Komponenten angegeben wird, welche Spracherkennungsrate jeweils mit den Merkmalsvektoren mit den jeweils berücksichtigten Merkmalsvektor-Komponenten erzielbar ist,
- einer Erkennungsraten-Informationen-Ermittlungseinheit zum Ermitteln der Erkennungsraten-Information,
- einer Informationsgehalt-Ermittlungseinheit, zum Ermitteln des Informationsgehalts für Merkmalsvektor-Komponenten eines Merkmalsvektors in dem Spracherkennungssystem,
- einer Merkmalsvektor-Komponenten-Auswahleinheit zum Auswählen von Merkmalsvektor-Komponenten, die im Rahmen der Spracherkennung zu berücksichtigen sind.

6. Spracherkennungssystem gemäß Anspruch 5,

bei dem die Spracherkennungseinheit eingerichtet ist zur sprecherunabhängigen Spracherkennung.

7. Spracherkennungssystem gemäß Ansprüche 5 oder 6, eingerichtet als ein Embedded System.

8. Steuereinrichtung zum Steuern eines technischen Systems mit einem Spracherkennungssystem gemäß einem der Ansprüche 5 bis 7,

wobei in dem elektronischen Wörterbuch die zum Steuern des technischen Systems vorgesehenen Steuerbefehle gespeichert sind.

9. Telekommunikationsgerät mit einer Steuereinrichtung gemäß
Anspruch 8.

Zusammenfassung

Verfahren zur rechnergestützten Spracherkennung, Spracherkennungssystem und Steuereinrichtung zum Steuern 5 eines technischen Systems

Es wird für eine ausgewählte Spracherkennungsanwendung bestimmt, welche Spracherkennungsrate erforderlich ist. Unter Verwendung einer gespeicherten Spracherkennungsrate-

- 10 Information wird ermittelt, welcher Informationsgehalt der Merkmalsvektor-Komponenten mindestens erforderlich ist, um die Spracherkennungsrate zu gewährleisten. Es wird die Anzahl der erforderlichen Merkmalsvektor-Komponenten ermittelt, die erforderlich ist, um den ermittelten Informationsgehalt
- 15 bereitzustellen, und die Spracherkennung wird ausgeführt unter Verwendung von Merkmalsvektoren mit der ermittelten benötigten Anzahl von Merkmalsvektor-Komponenten.

Signifikante Figur 4

Bezugszeichenliste

- 100 Spracherkennungssystem
- 101 Analoges Sprachsignal
- 102 Mikrofon
- 103 Analoges aufgenommenes Sprachsignal
- 104 Vorverarbeitung
- 105 Vorverarbeitetes Sprachsignal
- 106 Analog-/Digitalwandler
- 107 Digitales Signal
- 108 Merkmalsextraktionseinheit
- 109 Merkmalsvektor
- 110 Rechner
- 111 Eingangsschnittstelle
- 112 Mikroprozessor
- 113 Speicher
- 114 Ausgangsschnittstelle
- 115 Computerbus
- 116 Tastatur
- 117 Computermaus
- 118 Elektrische Leitung
- 119 Elektrische Leitung
- 120 Funkverbindung
- 121 Funkverbindung
- 122 Lautsprecher
- 123 Aktor

- 201 Erster Speicherteilbereich
- 202 Zweiter Speicherteilbereich
- 203 Dritter Speicherteilbereich
- 204 Vierter Speicherteilbereich
- 205 Fünfter Speicherteilbereich
- 206 Sechster Speicherteilbereich

- 300 Blockdiagramm
- 301 Sprachsignal
- 302 Segmentierungseinheit

303 LDA-Matrix-Berechnungseinheit

304 LDA-Matrix

305 Clustereinheit

306 Codebuch

307 Spracherkennungseinheit

308 Erkennungsrate

400 Ablaufdiagramm

401 Start

402 Segmentieren Sprachsignal

403 Berechnen LDA-Matrix

404 Auswahl Anzahl Merkmalsvektor-Komponenten

405 Clustern Merkmalsvektoren

406 Ermitteln Erkennungsrate

407 Prüfschritt

408 Ende

409 Reduktion Anzahl Merkmalsvektor-Komponenten der
Merkmalsvektoren

500 Funktionsdiagramm

501 Pseudoentropie

502 Erkennungsrate

503 Daten-Tupel

504 Schnittlinie

600 Ablaufdiagramm

601 Start

602 Auswählen Spracherkennungsanwendung

603 Segmentierung Sprachsignal

604 Berechnen LDA-Matrix

605 LDA-Matrix

606 Ermitteln erforderliche Pseudoentropie

607 Ermitteln Anzahl unnötiger Merkmalsvektor-Komponenten

608 Clustering

609 Spracherkennungsanwendungs-abhängiges Codebuch

610 Sprecherunabhängige Spracherkennung

611 Ende

FIG 1

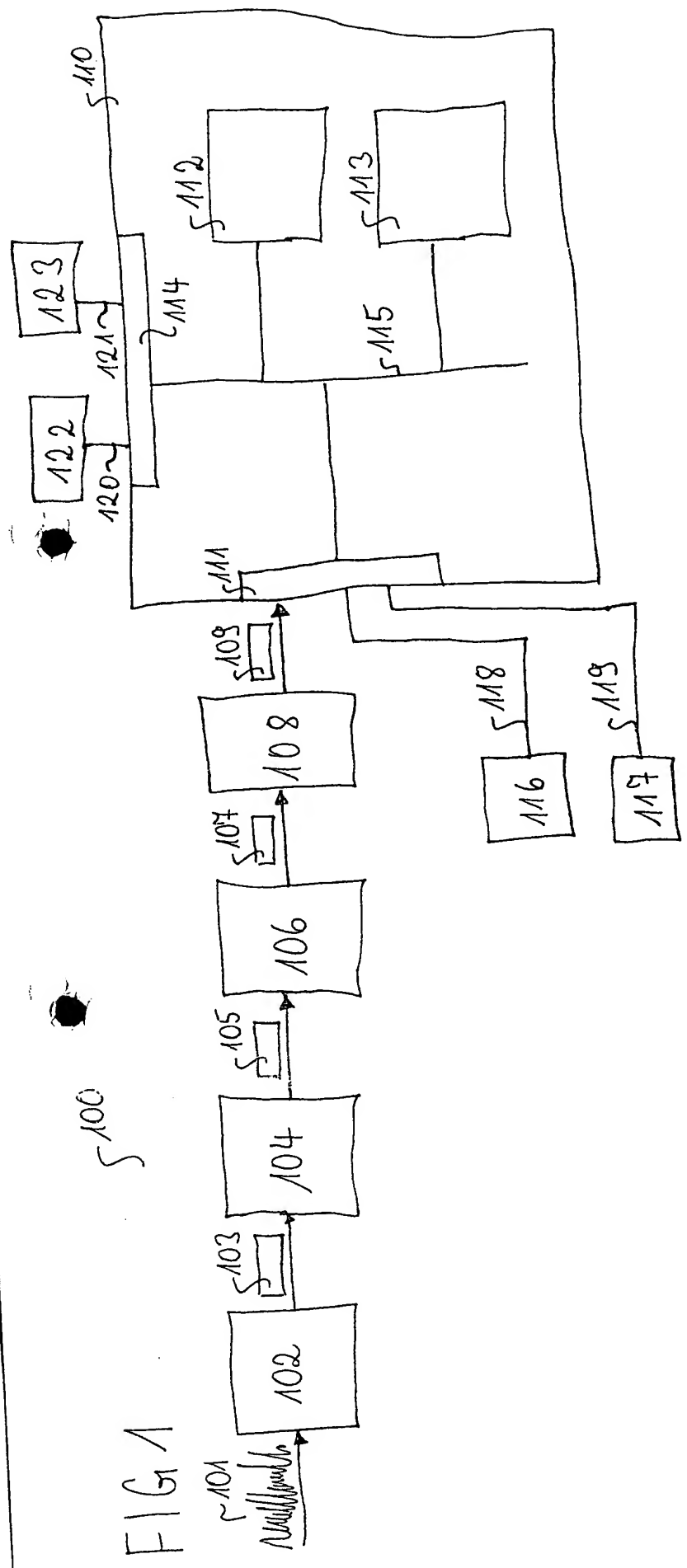


FIG 2

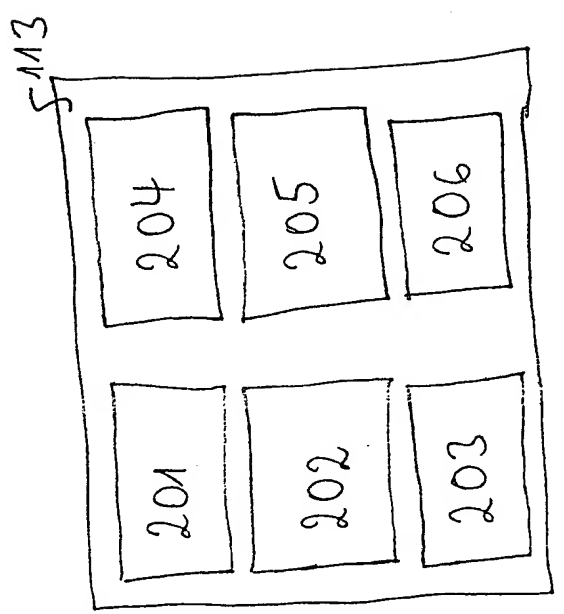


FIG 3

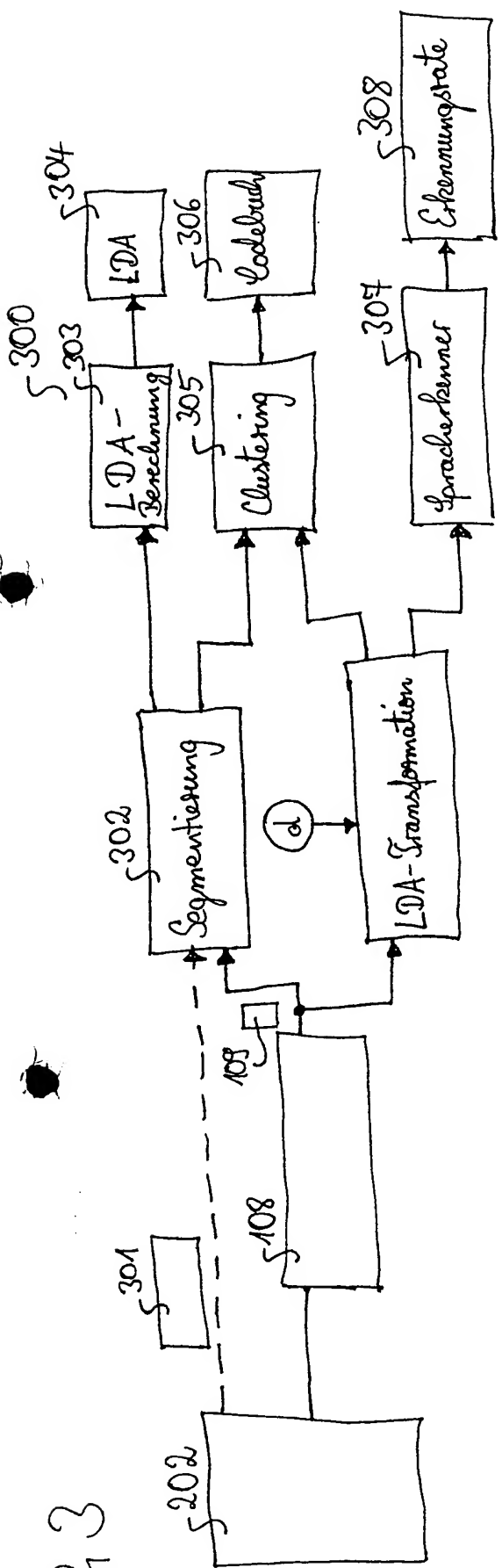
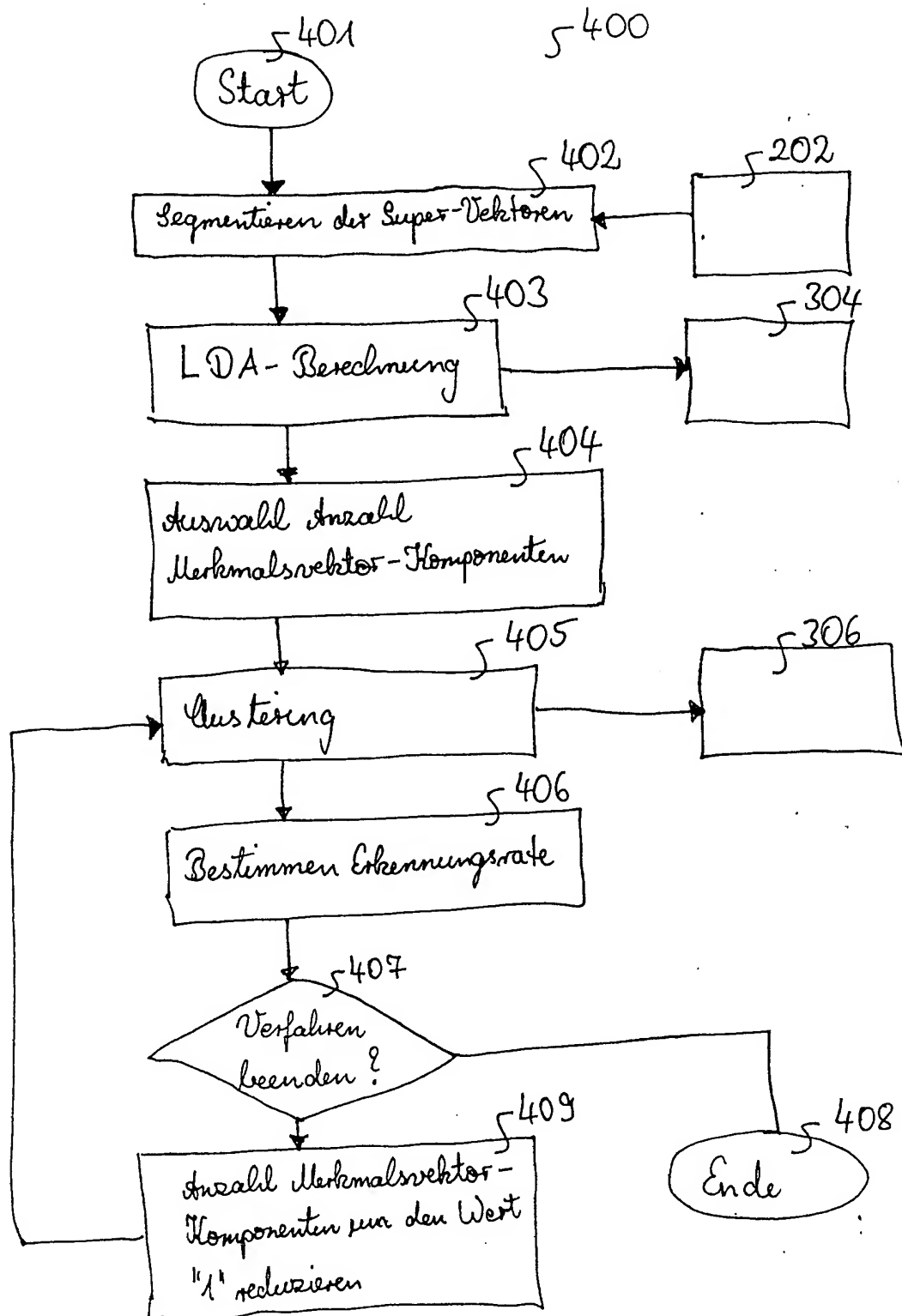


FIG 4



415

FIG. 5

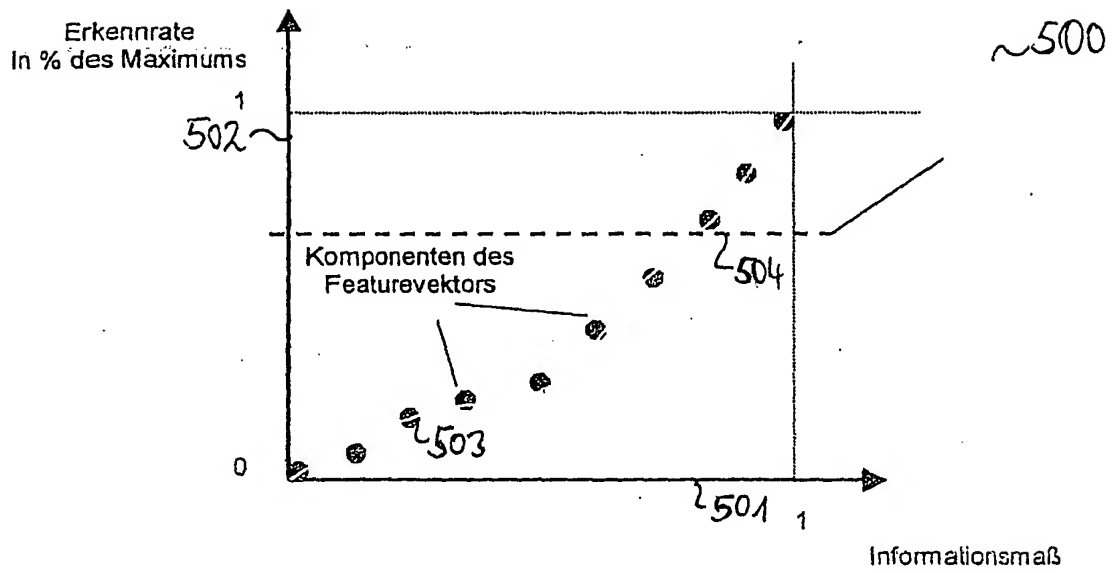


FIG 6

